

BEST AVAILABLE COPY

PCT/JP2004/008087

日 本 国 特 許 庁
JAPAN PATENT OFFICE

03.06.2004

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 4 年 4 月 9 日

出 願 番 号
Application Number: 特 願 2 0 0 4 - 1 4 2 9 0 7
[ST. 10/C]: [J P 2 0 0 4 - 1 4 2 9 0 7]

REC'D 22 JUL 2004

WIPO PCT

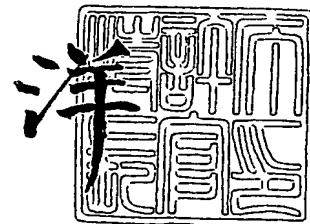
出 願 人
Applicant(s): 株式会社ケンウッド

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2 0 0 4 年 7 月 8 日

特許庁長官
Commissioner,
Japan Patent Office

小 川



出証番号 出証特 2 0 0 4 - 3 0 5 9 0 9 9

【書類名】 特許願
【整理番号】 P07-975772
【提出日】 平成16年 4月 9日
【あて先】 特許庁長官 殿
【国際特許分類】 G10L 13/06
G10L 13/08
G10L 11/04

【発明者】
【住所又は居所】 東京都八王子市石川町 2 9 6 7 - 3 株式会社ケンウッド内
【氏名】 佐藤 寧

【特許出願人】
【識別番号】 000003595
【氏名又は名称】 株式会社ケンウッド

【代理人】
【識別番号】 100095407
【弁理士】
【氏名又は名称】 木村 満

【先の出願に基づく優先権主張】
【出願番号】 特願2003-160657
【出願日】 平成15年 6月 5日

【手数料の表示】
【予納台帳番号】 038380
【納付金額】 16,000円

【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【包括委任状番号】 9903184

【書類名】 特許請求の範囲**【請求項 1】**

音片を表す音片データを複数記憶する音片記憶手段と、
文章を表す文章情報を入力し、
各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、
前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、
前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、
より構成されることを特徴とする音声合成装置。

【請求項 2】

音片を表す音片データを複数記憶する音片記憶手段と、
文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、
各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、
前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音片の波形を表す音声データを合成する欠落部分合成手段と、
前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、
より構成されることを特徴とする音声合成装置。

【請求項 3】

前記選択手段は、韻律が韻律予測結果に前記所定の条件下で合致しない音片データを、選択の対象から除外する、
ことを特徴とする請求項 2 に記載の音声合成装置。

【請求項 4】

前記欠落部分合成手段は、
音素を表し、又は、音素を構成する素片を表すデータを複数記憶する記憶手段と、
前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得して互いに結合することにより、当該音声の波形を表す音声データを合成する合成手段と、を備える、
ことを特徴とする請求項 2 又は 3 に記載の音声合成装置。

【請求項 5】

前記欠落部分合成手段は、前記選択手段が音片データを選択できなかった前記音声の韻律を予測する欠落部分韻律予測手段を備え、
前記合成手段は、前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得し、取得したデータを、当該データが表す音素又は素片が、前記欠落部分韻律予測手段による韻律の予測結果に合致するように変換して、変換されたデータを互いに結合することにより、当該音声の波形を表す音声データを合成する、
ことを特徴とする請求項 4 に記載の音声合成装置。

【請求項 6】

前記欠落部分合成手段は、前記韻律予測手段が予測した韻律に基づいて、前記選択手段が音片データを選択できなかった音声について、当該音片の波形を表す音声データを合成する、

ことを特徴とする請求項 2、3 又は 4 に記載の音声合成装置。

【請求項 7】

前記音片記憶手段は、音片データが表す音片のピッチの時間変化を表す韻律データを、当該音片データに対応付けて記憶しており、

前記選択手段は、各前記音片データのうちから、前記文章を構成する音声と読みが共通しており、且つ、対応付けられている韻律データが表すピッチの時間変化が韻律の予測結果に最も近い音片データを選択する、

ことを特徴とする請求項 2 乃至 6 のいずれか 1 項に記載の音声合成装置。

【請求項 8】

前記合成音声を発声するスピードの条件を指定する発声スピードデータを取得し、前記合成音声を表すデータを構成する音片データ及び／又は音声データを、当該発声スピードデータが指定する条件を満たすスピードで発話される音声を表すように選択又は変換する発話スピード変換手段を備える、

ことを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の音声合成装置。

【請求項 9】

前記発話スピード変換手段は、前記合成音声を表すデータを構成する音片データ及び／又は音声データから素片を表す区間を除去し、又は、当該音片データ及び／又は音声データに素片を表す区間を追加することによって、当該音片データ及び／又は音声データを、前記発声スピードデータが指定する条件を満たすスピードで発話される音声を表すよう変換する、

ことを特徴とする請求項 8 に記載の音声合成装置。

【請求項 10】

前記音片記憶手段は、音片データの読みを表す表音データを、当該音片データに対応付けて記憶しており、

前記選択手段は、前記文章を構成する音声の読みに合致する読みを表す表音データが対応付けられている音片データを、当該音声と読みが共通する音片データとして扱う、

ことを特徴とする請求項 1 乃至 9 のいずれか 1 項に記載の音声合成装置。

【請求項 11】

音片を表す音片データを複数記憶し、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択し、

前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、

選択した音片データ及び合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする音声合成方法。

【請求項 12】

音片を表す音片データを複数記憶し、

文章を表す文章情報を入力して、当該文章を構成する音声の韻律を予測し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択し、

前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、

選択した音片データ及び合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする音声合成方法。

【請求項 13】

コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声に

ついて、当該音声の波形を表す音声データを合成する欠落部分合成手段と、
前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、
して機能させるためのプログラム。

【請求項 14】

コンピュータを、
音片を表す音片データを複数記憶する音片記憶手段と、
文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、
各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、
韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、
前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声に
ついて、当該音声の波形を表す音声データを合成する欠落部分合成手段と、
前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、
して機能させるためのプログラム。

【書類名】 明細書**【発明の名称】 音声合成装置、音声合成方法及びプログラム****【技術分野】****【0001】**

この発明は、音声合成装置、音声合成方法及びプログラムに関する。

【背景技術】**【0002】**

音声合成する手法として、録音編集方式と呼ばれる手法がある。録音編集方式は、駅の音声案内システムや、車載用のナビゲーション装置などに用いられている。

録音編集方式は、単語と、この単語を読み上げる音声を表す音声データとを対応付けておき、音声合成する対象の文章を単語に区切ってから、これらの単語に対応付けられた音声データを取得してつなぎ合わせる、という手法である（例えば、特許文献1参照）。

【特許文献1】 特開平10-49193号公報

【発明の開示】**【発明が解決しようとする課題】****【0003】**

しかし、音声データを単につなぎ合わせた場合、音声データ同士の境界では通常、音声のピッチ成分の周波数が不連続的に変化する、等の理由で、合成音声が不自然なものとなる。

この問題を解決する手法としては、同一の音素を互いに異なった韻律で読み上げる音声を表す複数の音声データを用意し、一方で音声合成する対象の文章に韻律予測を施して、予測結果に合致する音声データを選び出してつなぎ合わせる、という手法が考えられる。

しかし、音声データを音素毎に用意して録音編集方式により自然な合成音声を得ようとすると、音声データを記憶する記憶装置には膨大な記憶容量が必要となる。また、検索する対象のデータの量も膨大なものとなる。

【0004】

この発明は、上記実状に鑑みてなされたものであり、簡単な構成で高速に自然な合成音声を得るための音声合成装置、音声合成方法及びプログラムを提供することを目的とする。

【課題を解決するための手段】**【0005】**

上記目的を達成するため、この発明の第1の観点にかかる音声合成装置は、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、より構成されることを特徴とする。

【0006】

また、この発明の第2の観点にかかる音声合成装置は、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音片の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、より構成されることを特徴とする。

【0007】

前記選択手段は、韻律が韻律予測結果に前記所定の条件下で合致しない音片データを、選択の対象から除外するものであってもよい。

【0008】

前記欠落部分合成手段は、

音素を表し、又は、音素を構成する素片を表すデータを複数記憶する記憶手段と、

前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得して互いに結合することにより、当該音声の波形を表す音声データを合成する合成手段と、を備えるものであってもよい。

【0009】

前記欠落部分合成手段は、前記選択手段が音片データを選択できなかった前記音声の韻律を予測する欠落部分韻律予測手段を備えてもよく、

前記合成手段は、前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得し、取得したデータを、当該データが表す音素又は素片が、前記欠落部分韻律予測手段による韻律の予測結果に合致するように変換して、変換されたデータを互いに結合することにより、当該音声の波形を表す音声データを合成するものであってもよい。

【0010】

前記欠落部分合成手段は、前記韻律予測手段が予測した韻律に基づいて、前記選択手段が音片データを選択できなかった音声について、当該音片の波形を表す音声データを合成するものであってもよい。

【0011】

前記音片記憶手段は、音片データが表す音片のピッチの時間変化を表す韻律データを、当該音片データに対応付けて記憶していてもよく、

前記選択手段は、各前記音片データのうちから、前記文章を構成する音声と読みが共通しており、且つ、対応付けられている韻律データが表すピッチの時間変化が韻律の予測結果に最も近い音片データを選択するものであってもよい。

【0012】

前記音声合成装置は、前記合成音声を発声するスピードの条件を指定する発声スピードデータを取得し、前記合成音声を表すデータを構成する音片データ及び／又は音声データを、当該発声スピードデータが指定する条件を満たすスピードで発話される音声を表すように選択又は変換する発話スピード変換手段を備えるものであってもよい。

【0013】

前記発話スピード変換手段は、前記合成音声を表すデータを構成する音片データ及び／又は音声データから素片を表す区間を除去し、又は、当該音片データ及び／又は音声データに素片を表す区間を追加することによって、当該音片データ及び／又は音声データを、前記発声スピードデータが指定する条件を満たすスピードで発話される音声を表すよう変換するものであってもよい。

【0014】

前記音片記憶手段は、音片データの読みを表す表音データを、当該音片データに対応付けて記憶していてもよく、

前記選択手段は、前記文章を構成する音声の読みに合致する読みを表す表音データが対応付けられている音片データを、当該音声と読みが共通する音片データとして扱うものであってもよい。

【0015】

また、この発明の第3の観点にかかる音声合成方法は、

音片を表す音片データを複数記憶し、
文章を表す文章情報を入力し、
各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択し、
前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、
選択した音片データ及び合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、
ことを特徴とする。

【0016】

また、この発明の第4の観点にかかる音声合成方法は、
音片を表す音片データを複数記憶し、
文章を表す文章情報を入力して、当該文章を構成する音声の韻律を予測し、
各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択し、
前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、
選択した音片データ及び合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、
ことを特徴とする。

【0017】

また、この発明の第5の観点にかかるプログラムは、
コンピュータを、
音片を表す音片データを複数記憶する音片記憶手段と、
文章を表す文章情報を入力し、
各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、
前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、
前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、
して機能させるためのものであることを特徴とする。

【0018】

また、この発明の第6の観点にかかるプログラムは、
コンピュータを、
音片を表す音片データを複数記憶する音片記憶手段と、
文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、
各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、
前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、
前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、
して機能させるためのものであることを特徴とする。

【発明の効果】**【0019】**

以上説明したように、この発明によれば、簡単な構成で高速に自然な合成音声を得るための音声合成装置、音声合成方法及びプログラムが実現される。

【発明を実施するための最良の形態】

【0020】

以下、図面を参照して、この発明の実施の形態を説明する。

(第1の実施の形態)

図1は、この発明の第1の実施の形態に係る音声合成システムの構成を示す図である。図示するように、この音声合成システムは、本体ユニットM1と、音片登録ユニットRとにより構成されている。

【0021】

本体ユニットM1は、言語処理部1と、一般単語辞書2と、ユーザ単語辞書3と、規則合成処理部4と、音片編集部5と、検索部6と、音片データベース7と、伸長部8と、話速変換部9とにより構成されている。このうち、規則合成処理部4は、音響処理部41と、検索部42と、伸長部43と、波形データベース44とより構成されている。

【0022】

言語処理部1、音響処理部41、検索部42、伸長部43、音片編集部5、検索部6、伸長部8及び話速変換部9は、いずれも、CPU (Central Processing Unit) やDSP (Digital Signal Processor) 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

【0023】

なお、言語処理部1、音響処理部41、検索部42、伸長部43、音片編集部5、検索部6、伸長部8及び話速変換部9の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。従って、例えば、伸長部43の機能を行うプロセッサが伸長部8の機能を行ってもよいし、1個のプロセッサが音響処理部41、検索部42及び伸長部43の機能を兼ねて行ってもよい。

【0024】

一般単語辞書2は、PROM (Programmable Read Only Memory) やハードディスク装置等の不揮発性メモリより構成されている。一般単語辞書2には、表意文字 (例えば、漢字など) を含む単語等と、この単語等の読みを表す表音文字 (例えば、カナや発音記号など) とが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。

【0025】

ユーザ単語辞書3は、EEPROM (Electrically Erasable / Programmable Read Only Memory) やハードディスク装置等のデータ書き換え可能な不揮発性メモリと、この不揮発性メモリへのデータの書き込みを制御する制御回路とにより構成されている。なお、プロセッサがこの制御回路の機能を行ってもよく、言語処理部1、音響処理部41、検索部42、伸長部43、音片編集部5、検索部6、伸長部8及び話速変換部9の一部又は全部の機能を行うプロセッサがユーザ単語辞書3の制御回路の機能を行うようにしてもよい。

ユーザ単語辞書3は、表意文字を含む単語等と、この単語等の読みを表す表音文字とを、ユーザの操作に従って外部より取得し、互いに対応付けて記憶する。ユーザ単語辞書3には、一般単語辞書2に記憶されていない単語等とその読みを表す表音文字とが格納されていれば十分である。

【0026】

波形データベース44は、PROMやハードディスク装置等の不揮発性メモリより構成されている。波形データベース44には、表音文字と、この表音文字が表す単位音声の波形を表す波形データをエントロピー符号化して得られる圧縮波形データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。単位音声は、規則合成方式の手法で用いられる程度の短い音声であり、具体的には、音素や、VCV (Vowel-Consonant-Vowel) 音節などの単位で区切られる音声である。なお、エントロピー符号化される前の波形データは、例えば、PCM (Pulse Code Modulation) されたデジタル形式のデータからなっていればよい。

【0027】

音片データベース7は、PROMやハードディスク装置等の不揮発性メモリより構成されている。

音片データベース7には、例えば、図2に示すデータ構造を有するデータが記憶されている。すなわち、図示するように、音片データベース7に格納されているデータは、ヘッダ部HDR、インデックス部IDX、ディレクトリ部DIR及びデータ部DATの4種に分かれている。

【0028】

なお、音片データベース7へのデータの格納は、例えば、この音声合成システムの製造者によりあらかじめ行われ、及び／又は、音片登録ユニットRが後述する動作を行うことにより行われる。

【0029】

ヘッダ部HDRには、音片データベース7を識別するデータや、インデックス部IDX、ディレクトリ部DIR及びデータ部DATのデータ量、データの形式、著作権等の帰属などを示すデータが格納される。

【0030】

データ部DATには、音片の波形を表す音片データをエントロピー符号化して得られる圧縮音片データが格納されている。

なお、音片とは、音声のうち音素1個以上を含む連続した1区間をいい、通常は単語1個分又は複数個分の区間からなる。音片は接続詞を含む場合もある。

また、エントロピー符号化される前の音片データは、上述の圧縮波形データの生成のためエントロピー符号化される前の波形データと同じ形式のデータ（例えば、PCMされたデジタル形式のデータ）からなっていればよい。

【0031】

ディレクトリ部DIRには、個々の圧縮音声データについて、

- (A) この圧縮音片データが表す音片の読みを示す表音文字を表すデータ（音片読みデータ）、
 - (B) この圧縮音片データが格納されている記憶位置の先頭のアドレスを表すデータ、
 - (C) この圧縮音片データのデータ長を表すデータ、
 - (D) この圧縮音片データが表す音片の発声スピード（再生した場合の時間長）を表すデータ（スピード初期値データ）、
 - (E) この音片のピッチ成分の周波数の時間変化を表すデータ（ピッチ成分データ）、
- が、互いに対応付けられた形で格納されている。（なお、音片データベース7の記憶領域にはアドレスが付されているものとする。）

【0032】

なお、図2は、データ部DATに含まれるデータとして、読みが「サイタマ」である音片の波形を表す、データ量1410hバイトの圧縮音片データが、アドレス001A36A6hを先頭とする論理的な位置に格納されている場合を例示している。（なお、本明細書及び図面において、末尾に“h”を付した数字は16進数を表す。）

【0033】

なお、上述の(A)～(E)のデータの集合のうち少なくとも(A)のデータ（すなわち音片読みデータ）は、音片読みデータが表す表音文字に基づいて決められた順位に従ってソートされた状態で（例えば、表音文字がカナであれば、五十音順に従って、アドレス降順に並んだ状態で）、音片データベース7の記憶領域に格納されている。

また、上述のピッチ成分データは、例えば、図示するように、音片のピッチ成分の周波数を音片の先頭からの経過時間の1次関数で近似した場合における、この1次関数の切片 β 及び勾配 α の値を示すデータからなっていればよい。（勾配 α の単位は例えば[ヘルツ/秒]であればよく、切片 β の単位は例えば[ヘルツ]であればよい。）

また、ピッチ成分データには更に、圧縮音片データが表す音片が鼻濁音化されているか

否か、及び、無声化されているか否かを表す図示しないデータも含まれているものとする。

【0034】

インデックス部 I D X には、ディレクトリ部 D I R のデータのおおよその論理的位置を音片読みデータに基づいて特定するためのデータが格納されている。具体的には、例えば、音片読みデータがカナを表すものであるとして、カナ文字と、先頭 1 字がこのカナ文字であるような音片読みデータがどのような範囲のアドレスにあるかを示すデータ（ディレクトリアドレス）とが、互いに対応付けて格納されている。

【0035】

なお、一般単語辞書 2、ユーザ単語辞書 3、波形データベース 4 4 及び音片データベース 7 の一部又は全部の機能を単一の不揮発性メモリが行うようにしてもよい。

【0036】

音片登録ユニット R は、図示するように、収録音片データセット記憶部 1 0 と、音片データベース作成部 1 1 と、圧縮部 1 2 とにより構成されている。なお、音片登録ユニット R は音片データベース 7 とは着脱可能に接続されていてもよく、この場合は、音片データベース 7 に新たにデータを書き込むときを除いては、音片登録ユニット R を本体ユニット M 1 から切り離れた状態で本体ユニット M 1 に後述の動作を行わせてよい。

【0037】

収録音片データセット記憶部 1 0 は、ハードディスク装置等のデータ書き換え可能な不揮発性メモリより構成されている。

収録音片データセット記憶部 1 0 には、音片の読みを表す表音文字と、この音片を人が実際に発声したものを集音して得た波形を表す音片データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。なお、この音片データは、例えば、P C M されたデジタル形式のデータからなっていればよい。

【0038】

音片データベース作成部 1 1 及び圧縮部 1 2 は、C P U 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、このプログラムに従って後述する処理を行う。

【0039】

なお、音片データベース作成部 1 1 及び圧縮部 1 2 の一部又は全部の機能を単一のプロセッサが行うようにしてもよく、また、言語処理部 1、音響処理部 4 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサが音片データベース作成部 1 1 や圧縮部 1 2 の機能を更に行ってもよい。また、音片データベース作成部 1 1 や圧縮部 1 2 の機能を行うプロセッサが、収録音片データセット記憶部 1 0 の制御回路の機能を兼ねてもよい。

【0040】

音片データベース作成部 1 1 は、収録音片データセット記憶部 1 0 より、互いに対応付けられている表音文字及び音片データを読み出し、この音片データが表す音声のピッチ成分の周波数の時間変化と、発声スピードとを特定する。

発声スピードの特定は、例えば、この音片データのサンプル数を数えることにより特定すればよい。

【0041】

一方、ピッチ成分の周波数の時間変化は、例えば、この音片データにケプストラム解析を施すことにより特定すればよい。具体的には、例えば、音片データが表す波形を時間軸上で多数の小部分へと区切り、得られたそれぞれの小部分の強度を、元の値の対数（対数の底は任意）に実質的に等しい値へと変換し、値が変換されたこの小部分のスペクトル（すなわち、ケプストラム）を、高速フーリエ変換の手法（あるいは、離散変換をフーリエ変換した結果を表すデータを生成する他の任意の手法）により求める。そして、このケプストラムの極大値を与える周波数のうちの最小値を、この小部分におけるピッチ成分の周波数として特定する。

【0042】

なお、ピッチ成分の周波数の時間変化は、例えば、特開2003-108172号公報に開示された手法に従って音片データをピッチ波形データへと変換してから、このピッチ波形データに基づいて特定するようにすると良好な結果が期待できる。具体的には、音片データをフィルタリングしてピッチ信号を抽出し、抽出されたピッチ信号に基づいて、音片データが表す波形を単位ピッチ長の区間へと区切り、各区間について、ピッチ信号との相関関係に基づいて位相のずれを特定して各区間の位相を揃えることにより、音片データをピッチ波形信号へと変換すればよい。そして、得られたピッチ波形信号を音片データとして扱い、ケプストラム解析を行う等することにより、ピッチ成分の周波数の時間変化を特定すればよい。

【0043】

一方、音片データベース作成部11は、収録音片データセット記憶部10より読み出した音片データを圧縮部12に供給する。

圧縮部12は、音片データベース作成部11より供給された音片データをエントロピー符号化して圧縮音片データを作成し、音片データベース作成部11に返送する。

【0044】

音片データの発声スピード及びピッチ成分の周波数の時間変化を特定し、この音片データがエントロピー符号化され圧縮音片データとなって圧縮部12より返送されると、音片データベース作成部11は、この圧縮音片データを、データ部DATを構成するデータとして、音片データベース7の記憶領域に書き込む。

【0045】

また、音片データベース作成部11は、書き込んだ圧縮音片データが表す音片の読みを示すものとして収録音片データセット記憶部10より読み出した表音文字を、音片読みデータとして音片データベース7の記憶領域に書き込む。

また、書き込んだ圧縮音片データの、音片データベース7の記憶領域内での先頭のアドレスを特定し、このアドレスを上述の(B)のデータとして音片データベース7の記憶領域に書き込む。

また、この圧縮音片データのデータ長を特定し、特定したデータ長を、(C)のデータとして音片データベース7の記憶領域に書き込む。

また、この圧縮音片データが表す音片の発声スピード及びピッチ成分の周波数の時間変化を特定した結果を示すデータを生成し、スピード初期値データ及びピッチ成分データとして音片データベース7の記憶領域に書き込む。

【0046】

次に、この音声合成システムの動作を説明する。

まず、言語処理部1が、この音声合成システムに音声を合成させる対象としてユーザが用意した、表意文字を含む文章（フリーテキスト）を記述したフリーテキストデータを外部から取得したとして説明する。

【0047】

なお、言語処理部1がフリーテキストデータを取得する手法は任意であり、例えば、図示しないインターフェース回路を介して外部の装置やネットワークから取得してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体（例えば、フロッピー（登録商標）ディスクやCD-ROMなど）から、この記録媒体ドライブ装置を介して読み取ってもよい。

【0048】

また、言語処理部1の機能を行っているプロセッサが、自ら実行している他の処理で用いたテキストデータを、フリーテキストデータとして、言語処理部1の処理へと引き渡すようにしてもよい。

プロセッサが実行する当該他の処理としては、例えば、音声を表す音声データを取得し、この音声データに音声認識を施すことにより、この音声を表す語句を特定し、特定した語句に基づいて、この音声の発話者の要求の内容を特定して、特定した要求を満足させる

ために実行すべき処理を特定して実行するようなエージェント装置の機能をプロセッサに行わせるための処理などが考えられる。

【0049】

フリーテキストデータを取得すると、言語処理部1は、このフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書2やユーザ単語辞書3を検索することにより特定する。そして、この表意文字を、特定した表音文字へと置換する。そして、言語処理部1は、フリーテキスト内の表意文字がすべて表音文字へと置換した結果得られる表音文字列を、音響処理部41へと供給する。

【0050】

音響処理部41は、言語処理部1より表音文字列を供給されると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を検索するように、検索部42に指示する。

【0051】

検索部42は、この指示に応答して波形データベース44を検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する。そして、索出された圧縮波形データを伸長部43へと供給する。

【0052】

伸長部43は、検索部42より供給された圧縮波形データを、圧縮される前の波形データへと復元し、検索部42へと返送する。検索部42は、伸長部43より返送された波形データを、検索結果として音響処理部41へと供給する。

音響処理部41は、検索部42より供給された波形データを、言語処理部1より供給された表音文字列内での各表音文字の並びに従った順序で、音片編集部5へと供給する。

【0053】

音片編集部5は、音響処理部41より波形データを供給されると、この波形データを、供給された順序で互いに結合し、合成音声を表すデータ（合成音声データ）として出力する。フリーテキストデータに基づいて合成されたこの合成音声は、規則合成方式の手法により合成された音声に相当する。

【0054】

なお、音片編集部5が合成音声データを出力する手法は任意であり、例えば、図示しないD/A（Digital-to-Analog）変換器やスピーカを介して、この合成音声データが表す合成音声を再生するようにしてもよい。また、図示しないインターフェース回路を介して外部の装置やネットワークに送出してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体へ、この記録媒体ドライブ装置を介して書き込んでもよい。また、音片編集部5の機能を行っているプロセッサが、自ら実行している他の処理へと、合成音声データを引き渡すようにしてもよい。

【0055】

次に、音響処理部41が、外部より配信された、表音文字列を表すデータ（配信文字列データ）を取得したとする。（なお、音響処理部41が配信文字列データを取得する手法も任意であり、例えば、言語処理部1がフリーテキストデータを取得する手法と同様の手法で配信文字列データを取得すればよい。）

【0056】

この場合、音響処理部41は、配信文字列データが表す表音文字列を、言語処理部1より供給された表音文字列と同様に扱う。この結果、配信文字列データが表す表音文字列に含まれる表音文字に対応する圧縮波形データが検索部42により索出され、圧縮される前の波形データが伸長部43により復元される。復元された各波形データは音響処理部41を介して音片編集部5へと供給され、音片編集部5が、この波形データを、配信文字列データが表す表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する。配信文字列データに基づいて合成されたこの合成音声データも、規則合成方式の手法により合成された音声を表す。

【0057】

次に、音片編集部 5 が、定型メッセージデータ、発声スピードデータ、及び照合レベルデータを取得したとする。

なお、定型メッセージデータは、定型メッセージを表音文字列として表すデータであり、発声スピードデータは、定型メッセージデータが表す定型メッセージの発声スピードの指定値（この定型メッセージを発声する時間長の指定値）を示すデータである。照合レベルデータは、検索部 6 が行う後述の検索処理における検索条件を指定するデータであり、以下では「1」、「2」又は「3」のいずれかの値をとるものとし、「3」が最も厳格な検索条件を示すものとする。

【0058】

また、音片編集部 5 が定型メッセージデータや発声スピードデータや照合レベルデータを取得する手法は任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で定型メッセージデータや発声スピードデータや照合レベルデータを取得すればよい。

【0059】

定型メッセージデータ、発声スピードデータ、及び照合レベルデータが音片編集部 5 に供給されると、音片編集部 5 は、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部 6 に指示する。

【0060】

検索部 6 は、音片編集部 5 の指示に応答して音片データベース 7 を検索し、該当する圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとを索出し、索出された圧縮波形データを伸長部 4 3 へと供給する。複数の圧縮音片データが共通の表音文字ないし表音文字列に該当する場合も、該当する圧縮音片データすべてが、音声合成に用いられるデータの候補として索出される。一方、圧縮音片データを索出できなかった音片があった場合、検索部 6 は、該当する音片を識別するデータ（以下、欠落部分識別データと呼ぶ）を生成する。

【0061】

伸長部 4 3 は、検索部 6 より供給された圧縮音片データを、圧縮される前の音片データへと復元し、検索部 6 へと返送する。検索部 6 は、伸長部 4 3 より返送された音片データと、索出された音片読みデータ、スピード初期値データ及びピッチ成分データとを、検索結果として話速変換部 9 へと供給する。また、欠落部分識別データを生成した場合は、この欠落部分識別データも話速変換部 9 へと供給する。

【0062】

一方、音片編集部 5 は、話速変換部 9 に対し、話速変換部 9 に供給された音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致するようにすることを指示する。

【0063】

話速変換部 9 は、音片編集部 5 の指示に応答し、検索部 6 より供給された音片データを指示に合致するように変換して、音片編集部 5 に供給する。具体的には、例えば、検索部 6 より供給された音片データの元の時間長を、索出されたスピード初期値データに基づいて特定した上、この音片データをリサンプリングして、この音片データのサンプル数を、音片編集部 5 の指示したスピードに合致する時間長にすればよい。

【0064】

また、話速変換部 9 は、検索部 6 より供給された音片読みデータ及びピッチ成分データも音片編集部 5 に供給し、欠落部分識別データを検索部 6 より供給された場合は、更にこの欠落部分識別データも音片編集部 5 に供給する。

【0065】

なお、発声スピードデータが音片編集部 5 に供給されていない場合、音片編集部 5 は、話速変換部 9 に対し、話速変換部 9 に供給された音片データを変換せずに音片編集部 5 に供給するよう指示すればよく、話速変換部 9 は、この指示に応答し、検索部 6 より供給さ

れた音片データをそのまま音片編集部5に供給すればよい。

【0066】

音片編集部5は、話速変換部9より音片データ、音片読みデータ及びピッチ成分データを供給されると、供給された音片データのうちから、定型メッセージを構成する音片の波形に近似できる波形を表す音片データを、音片1個につき1個ずつ選択する。ただし、音片編集部5は、いかなる条件を満たす波形を定型メッセージの音片に近い波形とするかを、取得した照合レベルデータに従って設定する。

【0067】

具体的には、まず、音片編集部5は、定型メッセージデータが表す定型メッセージに、例えば「藤崎モデル」や「ToBI (Tone and Break Indices)」等の韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律（アクセント、イントネーション、強勢、音素の時間長など）を予測する。

【0068】

次に、音片編集部5は、例えば、

(1) 照合レベルデータの値が「1」である場合は、話速変換部9より供給された音片データ（すなわち、定型メッセージ内の音片と読みが合致する音片データ）をすべて、定型メッセージ内の音片の波形に近いものとして選択する。

【0069】

(2) 照合レベルデータの値が「2」である場合は、(1)の条件（つまり、読みを表す表音文字の合致という条件）を満たし、更に、音片データのピッチ成分の周波数の時間変化を表すピッチ成分データの内容と定型メッセージに含まれる音片のアクセント（いわゆる韻律）の予測結果との間に所定量以上の強い相関がある場合（例えば、アクセントの位置の時間差が所定量以下である場合）に限り、この音片データが定型メッセージ内の音片の波形に近いものとして選択する。なお、定型メッセージ内の音片のアクセントの予測結果は、定型メッセージの韻律の予測結果より特定できるものであり、音片編集部5は、例えば、ピッチ成分の周波数が最も高いと予測されている位置をアクセントの予測位置であると解釈すればよい。一方、音片データが表す音片のアクセントの位置については、例えば、ピッチ成分の周波数が最も高い位置を上述のピッチ成分データに基づいて特定し、この位置をアクセントの位置であると解釈すればよい。また、韻律予測は、文章全体に対して行ってもよいし、文章を所定の単位に分割し、それぞれの単位に対して行ってもよい。

【0070】

(3) 照合レベルデータの値が「3」である場合は、(2)の条件（つまり、読みを表す表音文字及びアクセントの合致という条件）を満たし、更に、音片データが表す音声の鼻濁音化や無声化の有無が、定型メッセージの韻律の予測結果に合致している場合に限り、この音片データが定型メッセージ内の音片の波形に近いものとして選択する。音片編集部5は、音片データが表す音声の鼻濁音化や無声化の有無を、話速変換部9より供給されたピッチ成分データに基づいて判別すればよい。

【0071】

なお、音片編集部5は、自ら設定した条件に合致する音片データが1個の音片につき複数あった場合は、これら複数の音片データを、設定した条件より厳格な条件に従って1個に絞り込むものとする。

【0072】

具体的には、例えば、設定した条件が照合レベルデータの値「1」に相当するものであって、該当する音片データが複数あった場合は、照合レベルデータの値「2」に相当する検索条件にも合致するものを選択し、なお複数の音片データが選択された場合は、選択結果のうちから照合レベルデータの値「3」に相当する検索条件にも合致するものを更に選択する、等の操作を行う。照合レベルデータの値「3」に相当する検索条件で絞り込んでなお複数の音片データが残る場合は、残ったものを任意の基準で1個に絞り込めばよい。

【0073】

一方、音片編集部 5 は、話速変換部 9 より欠落部分識別データも供給されている場合には、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部 4 1 に供給し、この音片の波形を合成するよう指示する。

【0074】

指示を受けた音響処理部 4 1 は、音片編集部 5 より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が示す音声の波形を表す圧縮波形データが検索部 4 2 により索出され、この圧縮波形データが伸長部 4 3 により元の波形データへと復元され、検索部 4 2 を介して音響処理部 4 1 へと供給される。音響処理部 4 1 は、この波形データを音片編集部 5 へと供給する。

【0075】

音片編集部 5 は、音響処理部 4 1 より波形データを返送されると、この波形データと、話速変換部 9 より供給された音片データのうち音片編集部 5 が選択したものとを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

【0076】

なお、話速変換部 9 より供給されたデータに欠落部分識別データが含まれていない場合は、音響処理部 4 1 に波形の合成を指示することなく直ちに、音片編集部 5 が選択した音片データを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

【0077】

以上説明した、この発明の第 1 の実施の形態の音声合成システムでは、音素より大きな単位であり得る音片の波形を表す音片データが、韻律の予測結果に基づいて、録音編集方式により自然につながり合わせられ、定型メッセージを読み上げる音声合成される。音片データベース 7 の記憶容量は、音素毎に波形を記憶する場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追従できる。

【0078】

なお、この音声合成システムの構成は上述のものに限られない。

例えば、波形データや音片データは PCM 形式のデータである必要はなく、データ形式は任意である。

また、波形データベース 4 4 や音片データベース 7 は波形データや音片データを必ずしもデータ圧縮された状態で記憶している必要はない。波形データベース 4 4 や音片データベース 7 が波形データや音片データをデータ圧縮されていない状態で記憶している場合、本体ユニット M 1 は伸長部 4 3 を備えている必要はない。

【0079】

また、波形データベース 4 4 は、必ずしも単位音声を個々に分解された形で記憶している必要はなく、例えば、複数の単位音声からなる音声の波形と、この波形内で個々の単位音声占める位置を識別するデータとを記憶するようにしてもよい。またこの場合、音片データベース 7 が波形データベース 4 4 の機能を行ってもよい。つまり、波形データベース 4 4 内には、音片データベース 7 と同様の形式で一連の音声データが連なって記憶されていてもよく、この場合は、波形データベースとして利用するために、音声データ内の各音素毎に、表音文字やピッチ情報等が関連づけて記憶されているものとする。

【0080】

また、音片データベース作成部 1 1 は、図示しない記録媒体ドライブ装置にセットされた記録媒体から、この記録媒体ドライブ装置を介して、音片データベース 7 に追加する新たな圧縮音片データの材料となる音片データや表音文字列を読み取ってもよい。

また、音片登録ユニット R は、必ずしも収録音片データセット記憶部 1 0 を備えている必要はない。

【0081】

また、ピッチ成分データは音片データが表す音片のピッチ長の時間変化を表すデータで

あってもよい。この場合、音片編集部5は、ピッチ長が最も短い位置（つまり、周波数が高かったり低い位置）をピッチ成分データに基づいて特定し、この位置をアクセントの位置であると解釈すればよい。

【0082】

また、音片編集部5は、特定の音片の韻律を表す韻律登録データをあらかじめ記憶し、定型メッセージにこの特定の音片が含まれている場合は、この韻律登録データが表す韻律を、韻律予測の結果として扱うようにしてもよい。

また、音片編集部5は、過去の韻律予測の結果を韻律登録データとして新たに記憶するようにしてもよい。

【0083】

また、音片データベース作成部11は、マイクロフォン、増幅器、サンプリング回路、A/D (Analog-to-Digital) コンバータ及びPCMエンコーダなどを備えていてもよい。この場合、音片データベース作成部11は、収録音片データセット記憶部10より音片データを取得する代わりに、自己のマイクロフォンが集音した音声を表す音声信号を増幅し、サンプリングしてA/D変換した後、サンプリングされた音声信号にPCM変調を施すことにより、音片データを作成してもよい。

【0084】

また、音片編集部5は、音響処理部41より返送された波形データを話速変換部9に供給することにより、当該波形データが表す波形の時間長を、発声スピードデータが示すスピードに合致させるようにしてもよい。

【0085】

また、音片編集部5は、例えば、言語処理部1と共にフリーテキストデータを取得し、このフリーテキストデータが表すフリーテキストに含まれる音声（表音文字列）の少なくとも一部に合致する音片データを、定型メッセージの音片データの選択処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。

この場合、音響処理部41は、音片編集部5が選択した音片については、この音片の波形を表す波形データを検索部42に索出させなくてもよい。なお、音片編集部5は、音響処理部41が合成しなくてよい音片を音響処理部41に通知し、音響処理部41はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

【0086】

また、音片編集部5は、例えば、音響処理部41と共に配信文字列データを取得し、この配信文字列データが表す配信文字列に含まれる表音文字列を表す音片データを、定型メッセージの音片データの選択処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。この場合、音響処理部41は、音片編集部5が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部42に索出させなくてもよい。

【0087】

(第2の実施の形態)

次に、この発明の第2の実施の形態を説明する。図3は、この発明の第2の実施の形態に係る音声合成システムの構成を示す図である。図示するように、この音声合成システムも、第1の実施の形態におけるものと同様、本体ユニットM2と、音片登録ユニットRとにより構成されている。このうち、音片登録ユニットRの構成は、第1の実施の形態におけるものと実質的に同一の構成を有している。

【0088】

本体ユニットM2は、言語処理部1と、一般単語辞書2と、ユーザ単語辞書3と、規則合成処理部4と、音片編集部5と、検索部6と、音片データベース7と、伸長部8と、話速変換部9とにより構成されている。このうち、言語処理部1、一般単語辞書2、ユーザ単語辞書3及び音片データベース7は、第1の実施の形態におけるものと実質的に同一の構成を有している。

【0089】

言語処理部 1、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 は、いずれも、CPU や DSP 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。なお、言語処理部 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6 及び話速変換部 9 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。

【0090】

規則合成処理部 4 は、第 1 の実施の形態におけるものと同様、音響処理部 4 1 と、検索部 4 2 と、伸長部 4 3 と、波形データベース 4 4 とより構成されている。このうち、音響処理部 4 1、検索部 4 2 及び伸長部 4 3 はいずれも、CPU や DSP 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

【0091】

なお、音響処理部 4 1、検索部 4 2 及び伸長部 4 3 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。また、言語処理部 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサが、更に音響処理部 4 1、検索部 4 2 及び伸長部 4 3 の一部又は全部の機能を行うようにしてもよい。従って、例えば、伸長部 8 が規則合成処理部 4 の伸長部 4 3 の機能を兼ねて行うようにしてもよい。

【0092】

波形データベース 4 4 は、PROM やハードディスク装置等の不揮発性メモリより構成されている。波形データベース 4 4 には、表音文字と、この表音文字が表す音素を構成する素片（すなわち、1 個の音素を構成する音声の波形 1 サイクル分（又はその他所定数のサイクル分）の音声）を表す素片波形データをエントロピー符号化して得られる圧縮波形データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。なお、エントロピー符号化される前の素片波形データは、例えば、PCM されたデジタル形式のデータからなっていればよい。

【0093】

音片編集部 5 は、一致音片決定部 5 1 と、韻律予測部 5 2 と、出力合成部 5 3 とより構成されている。一致音片決定部 5 1、韻律予測部 5 2 及び出力合成部 5 3 はいずれも、CPU や DSP 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

なお、一致音片決定部 5 1、韻律予測部 5 2 及び出力合成部 5 3 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。また、言語処理部 1、音響処理部 4 1、検索部 4 2、伸長部 4 3、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサが、更に一致音片決定部 5 1、韻律予測部 5 2 及び出力合成部 5 3 の一部又は全部の機能を行うようにしてもよい。従って、例えば、出力合成部 5 3 の機能を行うプロセッサが話速変換部 9 の機能を行うようにしてもよい。

【0094】

次に、図 3 の音声合成システムの動作を説明する。

まず、言語処理部 1 が、第 1 の実施の形態におけるものと実質的に同一のフリーテキストデータを外部から取得したとする。この場合、言語処理部 1 は、第 1 の実施の形態における処理と実質的に同一の処理を行うことにより、このフリーテキストに含まれる表意文字を表音文字へと置換する。そして、置換を行った結果得られた表音文字列を、規則合成処理部 4 の音響処理部 4 1 に供給する。

【0095】

音響処理部 4 1 は、言語処理部 1 より表音文字列を供給されると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す音素を構成する素片の波形を検索するよう、検索部 4 2 に指示する。また、音響処理部 4 1 は、この表音文字列を、音片編集部 5 の韻律予測部 5 2 に供給する。

【0096】

検索部42は、この指示に応答して波形データベース44を検索し、この指示の内容に合致する圧縮波形データを索出する。そして、索出された圧縮波形データを伸長部43へと供給する。

【0097】

伸長部43は、検索部42より供給された圧縮波形データを、圧縮される前の素片波形データへと復元し、検索部42へと返送する。検索部42は、伸長部43より返送された素片波形データを、検索結果として音響処理部41へと供給する。

【0098】

一方、音響処理部41より表音文字列を供給された韻律予測部52は、この表音文字列に、例えば第1の実施の形態で音片編集部5が行うものと同様の韻律予測の手法に基づいた解析を加えることにより、この表音文字列が表す音声の韻律の予測結果を表す韻律予測データを生成する。そして、この韻律予測データを、音響処理部41に供給する。

【0099】

音響処理部41は、検索部42より素片波形データを供給され、韻律予測部52より韻律予測データを供給されると、供給された素片波形データを用いて、言語処理部1が供給した表音文字列に含まれるそれぞれの表音文字が表す音声の波形を表す音声波形データを生成する。

【0100】

具体的には、音響処理部41は、例えば、検索部42より供給された各々の素片波形データが表す素片により構成されている音素の時間長を、韻律予測部52より供給された韻律予測データに基づいて特定する。そして、特定した音素の時間長を、当該素片波形データが表す素片の時間長で除した値に最も近い整数を求め、当該素片波形データを、求めた整数に等しい個数分相互に結合することにより、音声波形データを生成すればよい。

【0101】

なお、音響処理部41は、音声波形データが表す音声の時間長を韻律予測データに基づいて決定するのみならず、音声波形データを構成する素片波形データを加工して、音声波形データが表す音声の、当該韻律予測データが示す韻律に合致する強度やイントネーション等を有するようにしてもよい。

【0102】

そして、音響処理部41は、生成された音声波形データを、言語処理部1より供給された表音文字列内での各表音文字の並びに従った順序で、音片編集部5の出力合成部53へと供給する。

【0103】

出力合成部53は、音響処理部41より音声波形データを供給されると、この音声波形データを、音響処理部41より供給された順序で互いに結合し、合成音声データとして出力する。フリーテキストデータに基づいて合成されたこの合成音声は、規則合成方式の手法により合成された音声に相当する。

【0104】

なお、第1の実施の形態の音片編集部5と同様、出力合成部53が合成音声データを出力する手法も任意である。従って、例えば、図示しないD/A変換器やスピーカを介して、この合成音声データが表す合成音声を再生するようにしてもよい。また、図示しないインターフェース回路を介して外部の装置やネットワークに送出してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体へ、この記録媒体ドライブ装置を介して書き込んでもよい。また、出力合成部53の機能を行っているプロセッサが、自ら実行している他の処理へと、合成音声データを引き渡すようにしてもよい。

【0105】

次に、音響処理部41が、第1の実施の形態におけるものと実質的に同一の配信文字列データを取得したとする。（なお、音響処理部41が配信文字列データを取得する手法も任意であり、例えば、言語処理部1がフリーテキストデータを取得する手法と同様の手法

で配信文字列データを取得すればよい。)

【0106】

この場合、音響処理部41は、配信文字列データが表す表音文字列を、言語処理部1より供給された表音文字列と同様に扱う。この結果、配信文字列データが表す表音文字列に含まれる表音文字が表す音素を構成する素片を表す圧縮波形データが検索部42により索出され、圧縮される前の素片波形データが伸長部43により復元される。一方で、韻律予測部52により、配信文字列データが表す表音文字列に韻律予測の手法に基づいた解析が加えられ、この結果、この表音文字列が表す音声の韻律の予測結果を表す韻律予測データが生成される。そして音響処理部41が、配信文字列データが表す表音文字列に含まれるそれぞれの表音文字が表す音声の波形を表す音声波形データを、復元された各素片波形データと、韻律予測データとに基づいて生成し、出力合成部53は、生成された音声波形データを、配信文字列データが表す表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する。配信文字列データに基づいて合成されたこの合成音声データも、規則合成方式の手法により合成された音声を表す。

【0107】

次に、音片編集部5の一致音片決定部51が、第1の実施の形態におけるものと実質的に同一の定型メッセージデータ、発声スピードデータ、及び照合レベルデータを取得したとする。(なお、一致音片決定部51が定型メッセージデータや発声スピードデータや照合レベルデータを取得する手法は任意であり、例えば、言語処理部1がフリーテキストデータを取得する手法と同様の手法で定型メッセージデータや発声スピードデータや照合レベルデータを取得すればよい。)

【0108】

定型メッセージデータ、発声スピードデータ、及び照合レベルデータが一致音片決定部51に供給されると、一致音片決定部51は、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部6に指示する。

【0109】

検索部6は、一致音片決定部51の指示に応答して、第1の実施の形態の検索部6と同様に音片データベース7を検索し、該当する圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとをすべて索出し、索出された圧縮波形データを伸長部43へと供給する。一方、圧縮音片データを索出できなかった音片があった場合は、該当する音片を識別する欠落部分識別データを生成する。

【0110】

伸長部43は、検索部6より供給された圧縮音片データを、圧縮される前の音片データへと復元し、検索部6へと返送する。検索部6は、伸長部43より返送された音片データと、索出された音片読みデータ、スピード初期値データ及びピッチ成分データとを、検索結果として話速変換部9へと供給する。また、欠落部分識別データを生成した場合は、この欠落部分識別データも話速変換部9へと供給する。

【0111】

一方、一致音片決定部51は、話速変換部9に対し、話速変換部9に供給された音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致するようにすることを指示する。

【0112】

話速変換部9は、一致音片決定部51の指示に応答し、検索部6より供給された音片データを指示に合致するように変換して、一致音片決定部51に供給する。具体的には、例えば、検索部6より供給された音片データを個々の音素を表す区間へと区切り、得られたそれぞれの区間について、当該区間から、当該区間が表す音素を構成する素片を表す部分を特定して、特定された部分を(1個もしくは複数個)複製して当該区間内に挿入したり、又は、当該区間から当該部分を(1個もしくは複数個)除去することによって、当該区

間の長さを調整することにより、この音片データ全体のサンプル数を、一致音片決定部 51 の指示したスピードに合致する時間長にすればよい。なお、話速変換部 9 は、各区分について、素片を表す部分を挿入又は除去する個数を、各区分が表す音素相互間の時間長の比率が実質的に変化しないように決定すればよい。こうすることにより、音素同士を単に結合して合成する場合に比べて、音声のより細かい調整が可能になる。

【0113】

また、話速変換部 9 は、検索部 6 より供給された音片読みデータ及びピッチ成分データも一致音片決定部 51 に供給し、欠落部分識別データを検索部 6 より供給された場合は、更にこの欠落部分識別データも一致音片決定部 51 に供給する。

【0114】

なお、発声スピードデータが一致音片決定部 51 に供給されていない場合、一致音片決定部 51 は、話速変換部 9 に対し、話速変換部 9 に供給された音片データを変換せずに一致音片決定部 51 に供給するよう指示すればよく、話速変換部 9 は、この指示に応答し、検索部 6 より供給された音片データをそのまま一致音片決定部 51 に供給すればよい。また、話速変換部 9 に供給された音片データのサンプル数が、一致音片決定部 51 の指示したスピードに合致する時間長に既に合致している場合も、話速変換部 9 は、この音片データを変換せずそのまま一致音片決定部 51 に供給すればよい。

【0115】

一致音片決定部 51 は、話速変換部 9 より音片データ、音片読みデータ及びピッチ成分データを供給されると、第 1 の実施の形態の音片編集部 5 と同様、照合レベルデータの値に相当する条件に従って、自己に供給された音片データのうちから、定型メッセージを構成する音片の波形に近似できる波形を表す音片データを、音片 1 個につき 1 個ずつ選択する。

【0116】

ただし、一致音片決定部 51 は、話速変換部 9 より供給された音片データのうちから、照合レベルデータの値に相当する条件を満たす音片データを選択できない音片があった場合、該当する音片を、検索部 6 が圧縮音片データを索出できなかった音片（つまり、上述の欠落部分識別データが示す音片）とみなして扱うことを決定するものとする。

【0117】

そして、一致音片決定部 51 は、照合レベルデータの値に相当する条件を満たすものとして選択した音片データを、出力合成部 53 へと供給する。

また、一致音片決定部 51 は、話速変換部 9 より欠落部分識別データも供給されている場合、又は、照合レベルデータの値に相当する条件を満たす音片データを選択できなかった音片があった場合には、欠落部分識別データが示す音片（照合レベルデータの値に相当する条件を満たす音片データを選択できなかった音片を含む）の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部 41 に供給し、この音片の波形を合成するよう指示する。

【0118】

指示を受けた音響処理部 41 は、一致音片決定部 51 より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が表す音素を構成する素片を表す圧縮波形データが検索部 42 により索出され、圧縮される前の素片波形データが伸長部 43 により復元される。一方で、韻律予測部 52 により、この表音文字列が表す音片の韻律の予測結果を表す韻律予測データが生成される。そして音響処理部 41 が、この表音文字列に含まれるそれぞれの表音文字が表す音声の波形を表す音声波形データを、復元された各素片波形データと、韻律予測データとに基づいて生成し、生成された音声波形データを、出力合成部 53 へと供給する。

【0119】

なお、一致音片決定部 51 は、韻律予測部 52 が既に生成して一致音片決定部 51 に供給した韻律予測データのうち、欠落部分識別データが示す音片に相当する部分を音響処理部 41 に供給するにしてもよく、この場合、音響処理部 41 は、改めて韻律予測部 5

2に当該音片の韻律予測を行わせる必要はない。このようにすれば、音片等の細かい単位毎に韻律予測を行う場合に比べて、より自然な発話が可能になる。

【0120】

出力合成部53は、一致音片決定部51より音片データを供給され、音響処理部41より、素片波形データより生成された音声波形データを供給されると、供給されたそれぞれの音声波形データに含まれる素片波形データの個数を調整することにより、当該音声波形データが表す音声の時間長を、一致音片決定部51より供給された音片データが表す音片の発声スピードと整合するようにする。

【0121】

具体的には、出力合成部53は、例えば、一致音片決定部51より音片データに含まれる上述の各区間が表す音素の時間長が元の時間長に対して増減した比率を特定し、音響処理部41より供給された音声波形データが表す音素の時間長が当該比率で変化するように、各音声波形データ内の素片波形データの個数を増加あるいは減少させればよい。なお、出力合成部53は、当該比率を特定するため、例えば、一致音片決定部51が供給した音片データの生成に用いられた元の音片データを検索部6より取得し、これら2個の音片データ内で互いに同一の音素を表す区間を1個ずつ特定すればよい。そして、一致音片決定部51が供給した音片データ内で特定した区間内に含まれる素片の個数が、検索部6より取得した音片データ内で特定した区間内に含まれる素片の個数に対して増減した比率を、音素の時間長の増減の比率として特定するようにすればよい。なお、音声波形データが表す音素の時間長が、一致音片決定部51より供給された音片データが表す音片のスピードに既に整合している場合、出力合成部53は、音声波形データ内の素片波形データの個数を調整する必要はない。

【0122】

そして、出力合成部53は、素片波形データの個数の調整が完了した音声波形データと、一致音片決定部51より供給された音片データとを、定型メッセージデータが示す定型メッセージ内での各音片ないし音素の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

【0123】

なお、話速変換部9より供給されたデータに欠落部分識別データが含まれていない場合は、音響処理部41に波形の合成を指示することなく直ちに、音片編集部5が選択した音片データを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

【0124】

以上説明した、この発明の第2の実施の形態の音声合成システムでも、音素より大きな単位であり得る音片の波形を表す音片データが、韻律の予測結果に基づいて、録音編集方式により自然につなが合わせられ、定型メッセージを読み上げる音声合成される。

【0125】

一方、適切な音片データを選択することができなかった音片は、音素より小さな単位である素片を表す圧縮波形データを用いて、規則合成方式の手法に従って合成される。圧縮波形データが素片の波形を表すものであるため、波形データベース44の記憶容量は、圧縮波形データが音素の波形を表すものである場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追従できる。

【0126】

また、素片を用いて規則合成を行えば、音素を用いて規則合成を行う場合と異なり、音素の端の部分に現れる特殊な波形の影響を受けることなく音声合成を行うことができるため、少ない種類の素片で自然な音声を得ることができる。

すなわち、人が発声する音声では、先行する音素から後続の音素へと遷移する境界で、これらの音素双方の影響を受けた特殊な波形が現れることが知られており、一方、規則合成に用いられる音素は、採取した段階で既にその端部にこの特殊な波形を含んでいるため

、音素を用いて規則合成を行う場合は、音素間の境界の波形の様々なパターンを再現可能とするために膨大な種類の音素を用意するか、あるいは、音素間の境界の波形が自然な音声とは異なった合成音声を合成することで満足する必要がある。しかし、素片を用いて規則合成を行う場合は、音素の端部以外の部分から素片を採取するようにすれば、音素間の境界の特殊な波形の影響をあらかじめ排除することができる。このため、膨大な種類の素片を用意することを要せず、自然な音声を得ることができる。

【0127】

なお、この発明の第2の実施の形態の音声合成システムの構成も、上述のものに限られない。

例えば、素片波形データはPCM形式のデータである必要はなく、データ形式は任意である。また、波形データベース44は素片波形データや音片データを必ずしもデータ圧縮された状態で記憶している必要はない。波形データベース44が素片波形データをデータ圧縮されていない状態で記憶している場合、本体ユニットM2は伸長部43を備えている必要はない。

【0128】

また、波形データベース44は、必ずしも素片の波形を個々に分解された形で記憶している必要はなく、例えば、複数の素片からなる音声の波形と、この波形内で個々の素片が占める位置を識別するデータとを記憶するようにしてもよい。またこの場合、音片データベース7が波形データベース44の機能を行ってもよい。

【0129】

また、一致音片決定部51は、第1の実施の形態の音片編集部5と同様に韻律登録データをあらかじめ記憶し、定型メッセージにこの特定の音片が含まれている場合にこの韻律登録データが表す韻律を韻律予測の結果として扱うようにしてもよく、また、過去の韻律予測の結果を韻律登録データとして新たに記憶するようにしてもよい。

【0130】

また、一致音片決定部51は、第1の実施の形態の音片編集部5と同様にフリーテキストデータや配信文字列データを取得し、これらが表すフリーテキストや配信文字列に含まれる音片の波形に近い波形を表す音片データを、定型メッセージに含まれる音片の波形に近い波形を表す音片データを選択する処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。この場合、音響処理部41は、一致音片決定部51が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部42に索出させなくてもよく、また、一致音片決定部51は、音響処理部41が合成しなくてよい音片を音響処理部41に通知し、音響処理部41はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

【0131】

波形データベース44が記憶する圧縮波形データは、必ずしも素片を表すものである必要はなく、例えば、第1の実施の形態と同様、波形データベース44が記憶する表音文字が表す単位音声の波形を表す波形データ、あるいは当該波形データをエントロピー符号化して得られるデータであってもよい。

【0132】

また、波形データベース44は、素片の波形を表すデータと、音素の波形を表すデータとを、両方記憶していてもよい。この場合、音響処理部41は、配信文字列等に含まれる表音文字が表す音素のデータを検索部42に索出させ、該当する音素が索出されなかった表音文字について、当該表音文字が表す音素を構成する素片を表すデータを検索部42に索出させ、索出された、素片を表すデータを用いて、音素を表すデータを生成するようにしてもよい。

【0133】

また、話速変換部9が、音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる手法は任意である。従って、話速変換部9は、例えば第1の実施の形態における処理と同様に、検索部6より供給された音片データをリサンプリングして、

この音片データのサンプル数を、一致音片決定部 51 の指示した発話スピードに合致する時間長に相当する数へと増減させてもよい。

【0134】

また、本体ユニット M2 は必ずしも話速変換部 9 を備えている必要はない。本体ユニット M2 が話速変換部 9 を備えない場合、韻律予測部 52 が発話スピードを予測し、一致音片決定部 51 は、検索部 6 が取得した音片データのうち、所定の判別条件下で発話スピードが韻律予測部 52 による予測の結果に合致するものを選択し、一方、発話スピードが当該予測の結果に合致しないものを選択の対象から除外するものとしてもよい。なお、音片データベース 7 は、音片の読みが共通で発話スピードが互いに異なる複数の音片データを記憶していてもよい。

【0135】

また、出力合成部 53 が、音声波形データが表す音素の時間長を、音片データが表す音片の発話スピードと整合させる手法も任意である。従って、出力合成部 53 は、例えば、一致音片決定部 51 より音片データに含まれる各区間が表す音素の時間長が元の時間長に対して増減した比率を特定した上、音声波形データをリサンプリングして、音声波形データのサンプル数を、一致音片決定部 51 の指示した発話スピードと整合する時間長に相当する数へと増減させてもよい。

【0136】

また、発話スピードは音片毎に異なってもよい。(従って、発話スピードデータは、音片毎に異なる発話スピードを指定するものであってもよい。)そして、出力合成部 53 は、互いに発話スピードが異なる 2 個の音片の間に位置する各音声の音声波形データについては、当該 2 個の音片の発話スピードを補間(例えば、直線補間)することにより、当該 2 個の音片の間にあるこれらの音声の発話スピードを決定し、決定した発話スピードに合致するように、これらの音声を表す音声波形データを変換するようにしてもよい。

【0137】

また、出力合成部 53 は、音響処理部 41 より返送された音声波形データが、フリーテキストや配信文字列を読み上げる音声を構成する音声を表すものであっても、これらの音声波形データを変換して、これらの音声の時間長を、例えば一致音片決定部 51 に供給されている発話スピードデータが示すスピードに合致させるようにしてもよい。

【0138】

また、上述のシステムでは、例えば韻律予測部 52 が、文章全体に対して韻律予測(発話スピードの予測も含む)を行ってもよいし、所定の単位ごとに韻律予測を行ってもよい。また、文章全体に対して韻律予測を行った場合、読みが一致する音片があれば更に韻律が所定条件内で一致するか否かを判別し、一致していれば当該音片を採用するようにしてもよい。一致する音片が存在しなかった部分については、規則合成処理部 4 が素片を基に音声を生成するものとし、ただし、素片を基に合成する部分のピッチやスピードを、文章全体若しくは所定の単位ごとに行われた韻律予測の結果に基づいて調整するものとしてもよい。これによって、音片と、素片を基に生成する音声とを組み合わせる場合でも、自然な発話が行われる。

【0139】

また、言語処理部 1 に入力される文字列が表音文字列である場合、言語処理部 1 は、韻律予測とは別に公知の自然言語解析処理を行い、一致音片決定部 51 が、自然言語解析処理の結果に基づいて音片の選択を行ってもよい。これによって、単語(名詞や動詞等の品詞)毎に文字列を解釈した結果を用いて音片選択を行うことが可能になり、単に表音文字列と一致する音片を選択する場合に比べて自然な発話を行うことができる。

【0140】

以上、この発明の実施の形態を説明したが、この発明にかかる音声合成装置は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。

例えば、パーソナルコンピュータに上述の言語処理部 1、一般単語辞書 2、ユーザ単語辞書 3、音響処理部 41、検索部 42、伸長部 43、波形データベース 44、音片編集部

5、検索部6、音片データベース7、伸長部8及び話速変換部9の動作を実行させるためのプログラムを格納した記録媒体（CD-ROM、MO、フロッピー（登録商標）ディスク等）から該プログラムをインストールすることにより、上述の処理を実行する本体ユニットM1を構成することができる。

また、パーソナルコンピュータに上述の収録音片データセット記憶部10、音片データベース作成部11及び圧縮部12の動作を実行させるためのプログラムを格納した媒体から該プログラムをインストールすることにより、上述の処理を実行する音片登録ユニットRを構成することができる。

【0141】

そして、これらのプログラムを実行し本体ユニットM1や音片登録ユニットRとして機能するパーソナルコンピュータが、図1の音声合成システムの動作に相当する処理として、図4～図6に示す処理を行うものとする。

図4は、このパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

図5は、このパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

図6は、このパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

【0142】

すなわち、このパーソナルコンピュータが、外部より、上述のフリーテキストデータを取得すると（図4、ステップS101）、このフリーテキストデータが表すフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書2やユーザ単語辞書3を検索することにより特定し、この表意文字を、特定した表音文字へと置換する（ステップS102）。なお、このパーソナルコンピュータがフリーテキストデータを取得する手法は任意である。

【0143】

そして、このパーソナルコンピュータは、フリーテキスト内の表意文字をすべて表音文字へと置換した結果を表す表音文字列が得られると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース44より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップS103）。

【0144】

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップS104）、復元された波形データを、表音文字列内の各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する（ステップS105）。なお、このパーソナルコンピュータが合成音声データを出力する手法は任意である。

【0145】

また、このパーソナルコンピュータが、外部より、上述の配信文字列データを任意の手法で取得すると（図5、ステップS201）、この配信文字列データが表す表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース44より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップS202）。

【0146】

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップS203）、復元された波形データを、表音文字列内の各表音文字の並びに従った順序で互いに結合し、合成音声データとしてステップS105の処理と同様の処理により出力する（ステップS204）。

【0147】

一方、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発

声スピードデータを任意の手法により取得すると（図6、ステップS301）、まず、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出する（ステップS302）。

【0148】

また、ステップS302では、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データも索出する。なお、1個の音片につき複数の圧縮音片データが該当する場合は、該当する圧縮音片データすべてを索出する。一方、圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

【0149】

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の音片データへと復元する（ステップS303）。そして、復元された音片データを、上述の音片編集部5が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる（ステップS304）。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

【0150】

次に、このパーソナルコンピュータは、定型メッセージデータが表す定型メッセージに韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律を予測する（ステップS305）。そして、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形に最も近い波形を表す音片データを、上述の音片編集部5が行う処理と同様の処理を行うことにより、外部より取得した照合レベルデータが示す基準に従って、音片1個につき1個ずつ選択する（ステップS306）。

【0151】

具体的には、ステップS306でこのパーソナルコンピュータは、例えば、上述した（1）～（3）の条件に従って音片データを特定する。すなわち、照合レベルデータの値が「1」である場合は、定型メッセージ内の音片と読みが合致する音片データをすべて、定型メッセージ内の音片の波形を表しているとみなす。また、照合レベルデータの値が「2」である場合は、読みを表す表音文字が合致し、更に、音片データのピッチ成分の周波数の時間変化を表すピッチ成分データの内容が定型メッセージに含まれる音片のアクセントの予測結果に合致する場合に限り、この音片データが定型メッセージ内の音片の波形を表しているとみなす。また、照合レベルデータの値が「3」である場合は、読みを表す表音文字及びアクセントが合致し、更に、音片データが表す音声の鼻濁音化や無声化の有無が、定型メッセージの韻律の予測結果に合致している場合に限り、この音片データが定型メッセージ内の音片の波形を表しているとみなす。

なお、照合レベルデータが示す基準に合致する音片データが1個の音片につき複数あった場合は、これら複数の音片データを、設定した条件より厳格な条件に従って1個に絞り込むものとする。

【0152】

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップS202～S203の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す波形データを復元する（ステップS307）。

【0153】

そして、このパーソナルコンピュータは、復元した波形データと、ステップS306で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップS308）。

【0154】

また、例えば、パーソナルコンピュータに図3の言語処理部1、一般単語辞書2、ユーザ単語辞書3、音響処理部41、検索部42、伸長部43、波形データベース44、音片編集部5、検索部6、音片データベース7、伸長部8及び話速変換部9の動作を実行させるためのプログラムを格納した記録媒体から該プログラムをインストールすることにより、上述の処理を実行する本体ユニットM2を構成することもできる。

【0155】

そして、このプログラムを実行し本体ユニットM2として機能するパーソナルコンピュータが、図3の音声合成システムの動作に相当する処理として、図7～図9に示す処理を行うようにすることもできる。

図7は、本体ユニットM2の機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

図8は、本体ユニットM2の機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

図9は、本体ユニットM2の機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

【0156】

すなわち、このパーソナルコンピュータが、外部より、上述のフリーテキストデータを取得すると（図7、ステップS401）、このフリーテキストデータが表すフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書2やユーザ単語辞書3を検索することにより特定し、この表意文字を、特定した表音文字へと置換する（ステップS402）。なお、このパーソナルコンピュータがフリーテキストデータを取得する手法は任意である。

【0157】

そして、このパーソナルコンピュータは、フリーテキスト内の表意文字をすべて表音文字へと置換した結果を表す表音文字列が得られると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース44より検索し、表音文字列に含まれるそれぞれの表音文字が表す音素を構成する素片の波形を表す圧縮波形データを索出し（ステップS403）、索出された圧縮波形データを、圧縮される前の素片波形データへと復元する（ステップS404）。

【0158】

一方で、このパーソナルコンピュータは、フリーテキストデータに韻律予測の手法に基づいた解析を加えることにより、フリーテキストが表す音声の韻律を予測する（ステップS405）。そして、ステップS404で復元された素片波形データと、ステップS405における韻律の予測結果とに基づいて音声波形データを生成し（ステップS406）、得られた音声波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する（ステップS407）。なお、このパーソナルコンピュータが合成音声データを出力する手法は任意である。

【0159】

また、このパーソナルコンピュータが、外部より、上述の配信文字列データを任意の手法で取得すると（図8、ステップS501）、この配信文字列データが表す表音文字列に含まれるそれぞれの表音文字について、上述のステップS403～404と同様に、当該表音文字が表す音素を構成する素片の波形を表す圧縮波形データを索出する処理、及び、索出された圧縮波形データを素片波形データへと復元する処理を行う（ステップS502）。

【0160】

一方でこのパーソナルコンピュータは、配信文字列に韻律予測の手法に基づいた解析を加えることにより、配信文字列が表す音声の韻律を予測し（ステップS503）、ステップS502で復元された素片波形データと、ステップS503における韻律の予測結果とに基づいて音声波形データを生成し（ステップS504）、得られた音声波形データを、

表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとしてステップS407の処理と同様の処理により出力する(ステップS505)。

【0161】

一方、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると(図9、ステップS601)、まず、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出する(ステップS602)。

【0162】

また、ステップS602では、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データも索出する。なお、1個の音片につき複数の圧縮音片データが該当する場合は、該当する圧縮音片データすべてを索出する。一方、圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

【0163】

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の素片音片データへと復元する(ステップS603)。そして、復元された音片データを、上述の出力合成部53が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる(ステップS604)。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

【0164】

次に、このパーソナルコンピュータは、定型メッセージデータが表す定型メッセージに韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律を予測する(ステップS605)。そして、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形に最も近い波形を表す音片データを、上述の一致音片決定部51が行う処理と同様の処理を行うことにより、外部より取得した照合レベルデータが示す基準に従って、音片1個につき1個ずつ選択する(ステップS606)。

【0165】

具体的には、ステップS606でこのパーソナルコンピュータは、例えば、上述のステップ306の処理と同様の処理を行うことにより、上述した(1)～(3)の条件に従って音片データを特定する。なお、照合レベルデータが示す基準に合致する音片データが1個の音片につき複数あった場合は、これら複数の音片データを、設定した条件より厳格な条件に従って1個に絞り込むものとする。また、照合レベルデータの値に相当する条件を満たす音片データを選択できない音片があった場合は、該当する音片を、圧縮音片データを索出できなかった音片として扱うことと決定し、例えば欠落部分識別データを生成するものとする。

【0166】

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップS502～S504の処理と同様の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す音声波形データを生成する(ステップS607)。

ただし、ステップS607でこのパーソナルコンピュータは、ステップS503の処理に相当する処理を行う代わりに、ステップS605における韻律予測の結果を用いて音声波形データを生成するようにしてもよい。

【0167】

次に、このパーソナルコンピュータは、上述の出力合成部53が行う処理と同様の処理を行うことにより、ステップS607で生成された音声波形データに含まれる素片波形データの個数を調整し、当該音声波形データが表す音声の時間長を、ステップS606で選

択された音片データが表す音片の発声スピードと整合するようにする（ステップS608）。

【0168】

すなわち、ステップS608でこのパーソナルコンピュータは、例えば、ステップS606で選択された音片データに含まれる上述の各区間が表す音素の時間長が元の時間長に対して増減した比率を特定し、ステップS607で生成された音声波形データが表す音声の時間長が当該比率で変化するように、各音声波形データ内の素片波形データの個数を増加あるいは減少させればよい。なお、当該比率を特定するため、例えば、ステップS606で選択された音片データ（発声スピード変換後の音片データ）と、当該音片データがステップS604で変換を受ける前の元の音片データとの内で互いに同一の音声を表す区間を1個ずつ特定し、発声スピード変換後の音片データ内で特定した区間内に含まれる素片の個数が、元の音片データ内で特定した区間内に含まれる素片の個数に対して増減した比率を、音声の時間長の増減の比率として特定するようにすればよい。なお、音声波形データが表す音声の時間長が、発声スピード変換後の音片データが表す音片のスピードに既に整合している場合、このパーソナルコンピュータは音声波形データ内の素片波形データの個数を調整する必要はない。

【0169】

そして、このパーソナルコンピュータは、ステップS608の処理を経た音声波形データと、ステップS606で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップS609）。

【0170】

なお、パーソナルコンピュータに本体ユニットM1や本体ユニットM2や音片登録ユニットRの機能を行わせるプログラムは、例えば、通信回線の掲示板（BBS）にアップロードし、これを通信回線を介して配信してもよく、また、これらのプログラムを表す信号により搬送波を変調し、得られた変調波を伝送し、この変調波を受信した装置が変調波を復調してこれらのプログラムを復元するようにしてもよい。

そして、これらのプログラムを起動し、OSの制御下に、他のアプリケーションプログラムと同様に実行することにより、上述の処理を実行することができる。

【0171】

なお、OSが処理の一部を分担する場合、あるいは、OSが本願発明の1つの構成要素の一部を構成するような場合には、記録媒体には、その部分を除いたプログラムを格納してもよい。この場合も、この発明では、その記録媒体には、コンピュータが実行する各機能又はステップを実行するためのプログラムが格納されているものとする。

【図面の簡単な説明】

【0172】

【図1】この発明の第1の実施の形態に係る音声合成システムの構成を示すブロック図である。

【図2】音片データベースのデータ構造を模式的に示す図である。

【図3】この発明の第2の実施の形態に係る音声合成システムの構成を示すブロック図である。

【図4】この発明の第1の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

【図5】この発明の第1の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

【図6】この発明の第1の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

【図7】図3の本体ユニットの機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

【図8】図3の本体ユニットの機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

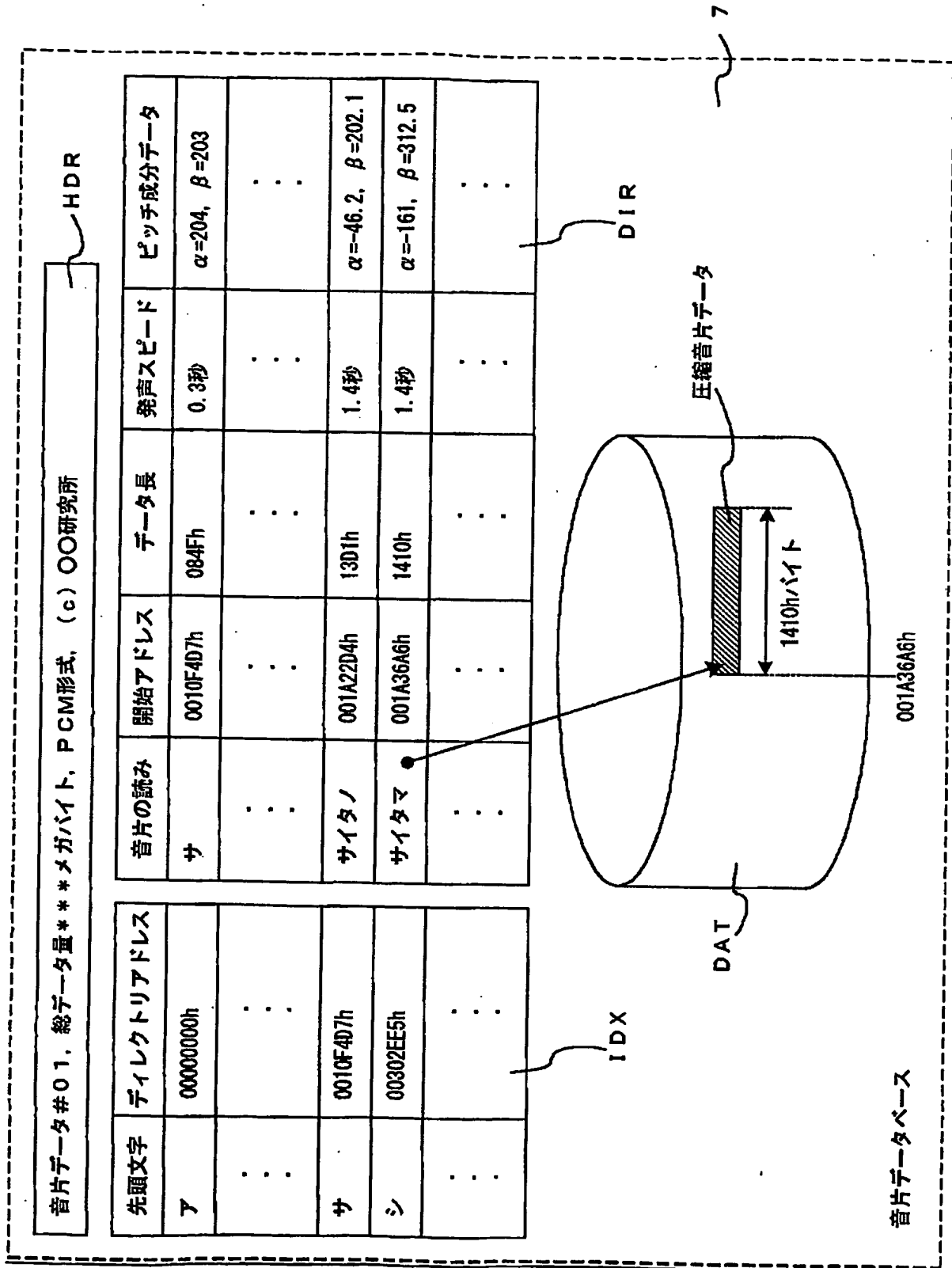
【図9】図3の本体ユニットの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

【符号の説明】

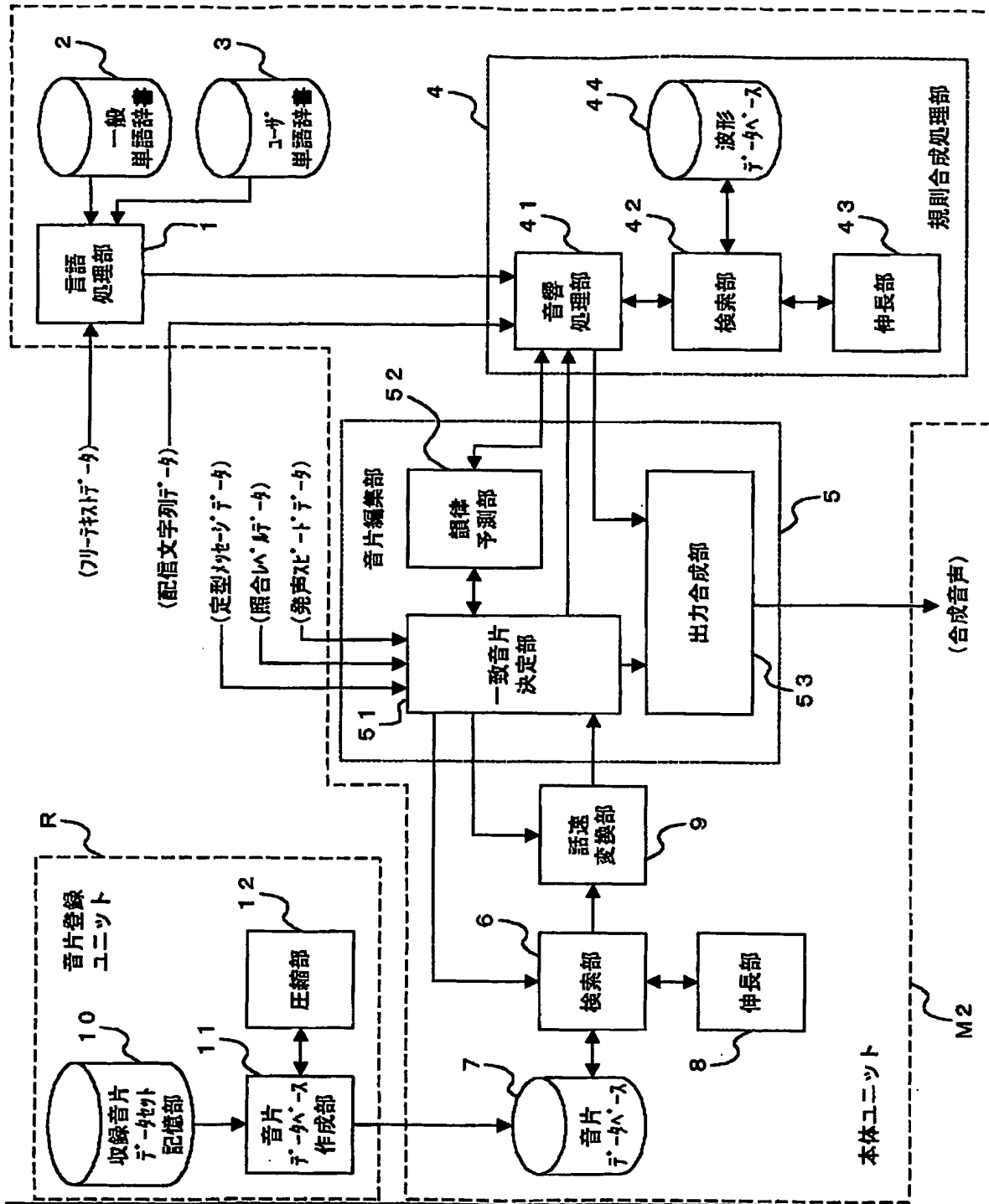
【0173】

M1, M2	本体ユニット
1	言語処理部
2	一般単語辞書
3	ユーザ単語辞書
41	音響処理部
42	検索部
43, 8	伸長部
44	波形データベース
5	音片編集部
51	一致音片決定部
52	韻律予測部
53	出力合成部
6	検索部
7	音片データベース
9	話速変換部
R	音片登録ユニット
10	収録音片データセット記憶部
11	音片データベース作成部
12	圧縮部
HDR	ヘッダ部
IDX	インデックス部
DIR	ディレクトリ部
DAT	データ部

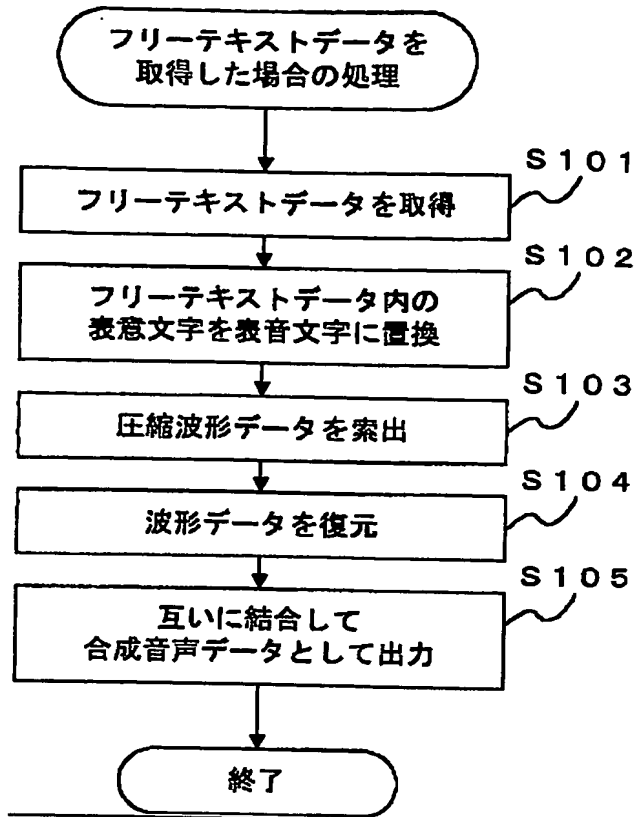
【図 2】



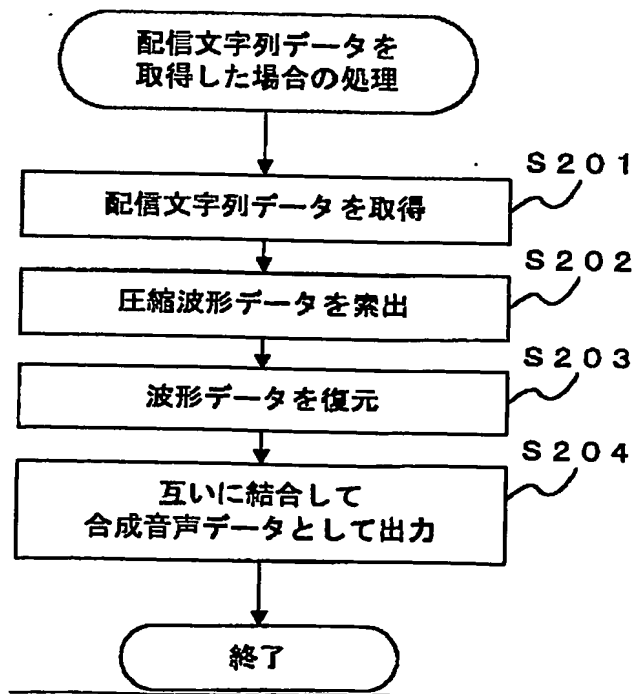
【図3】



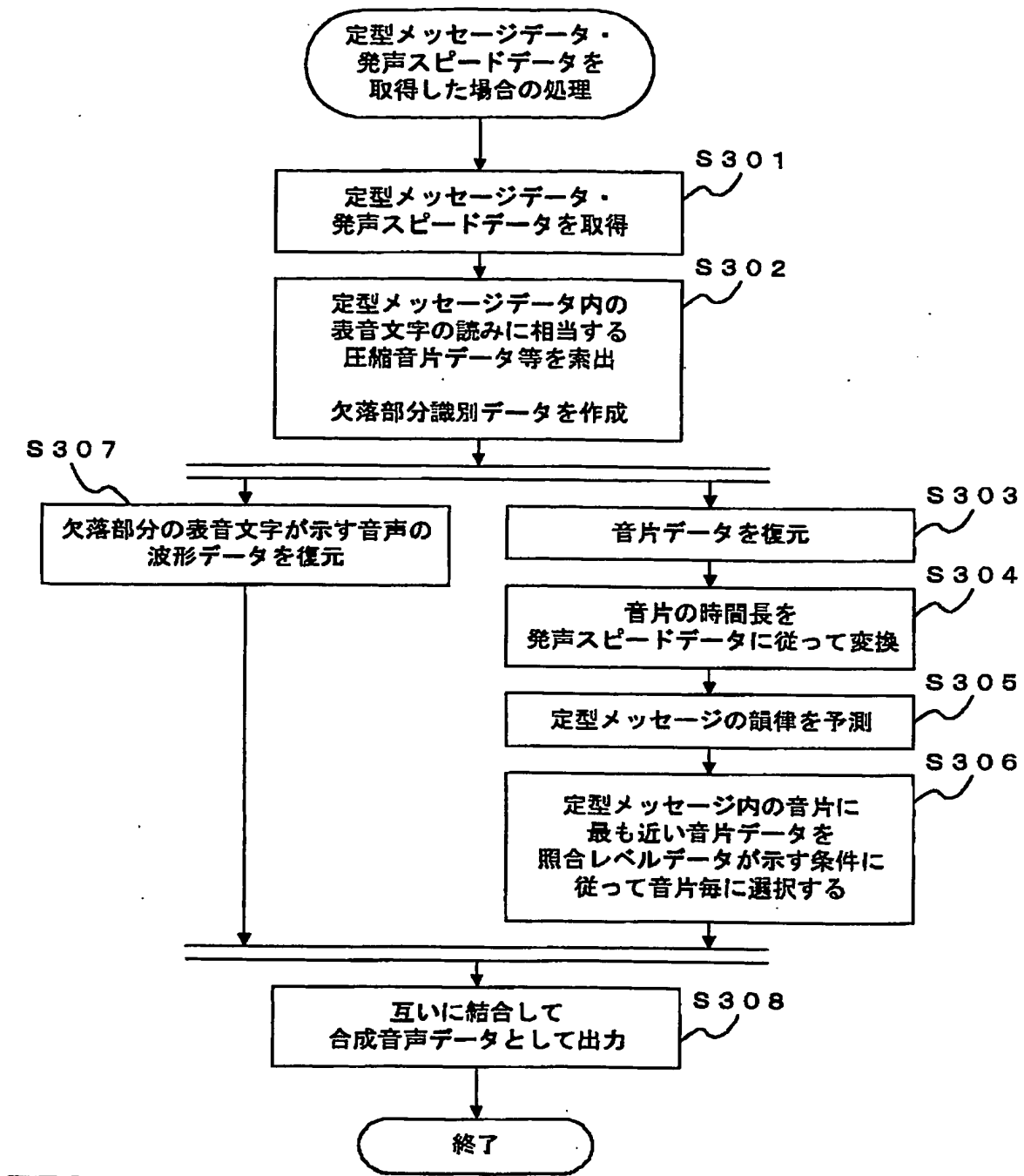
【図 4】



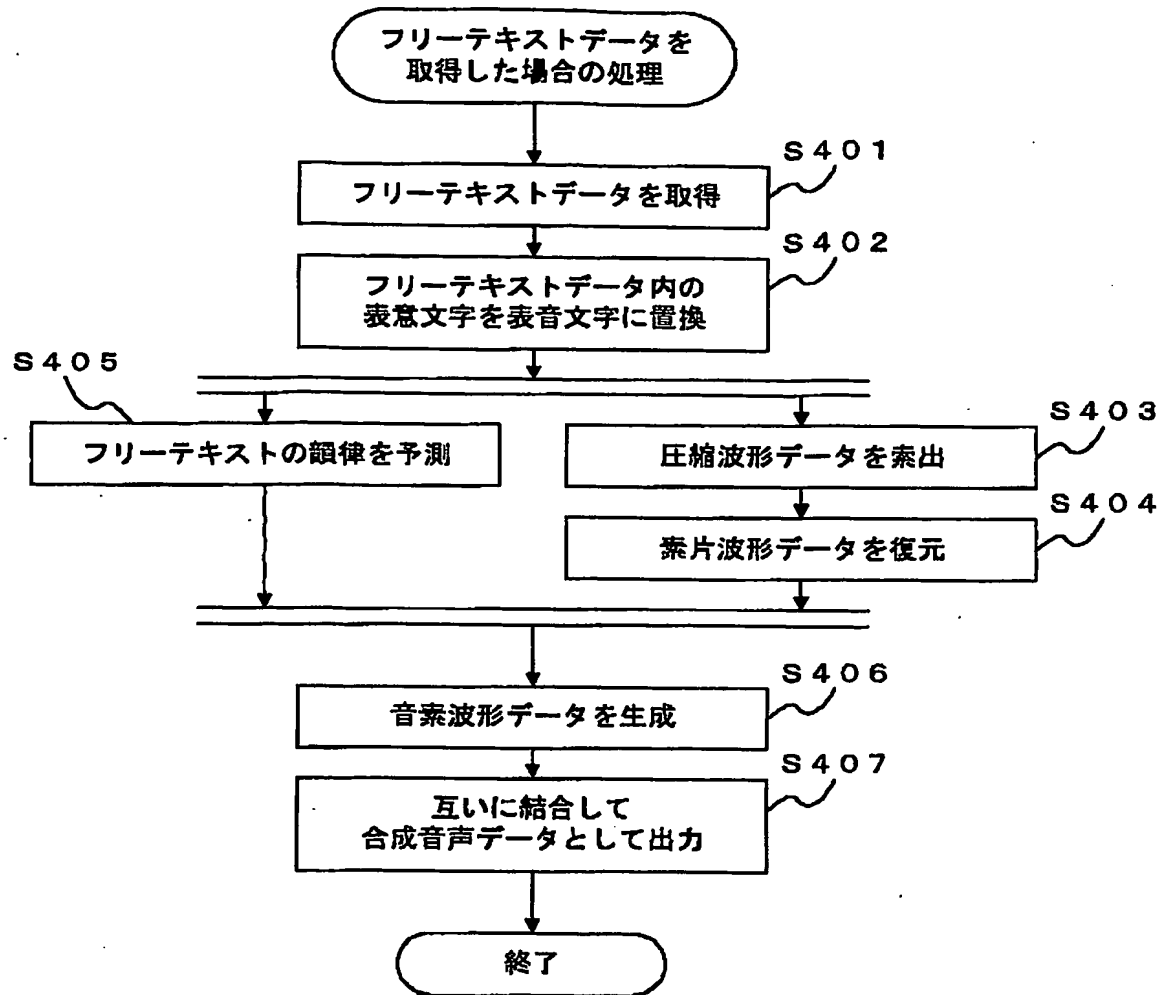
【図 5】



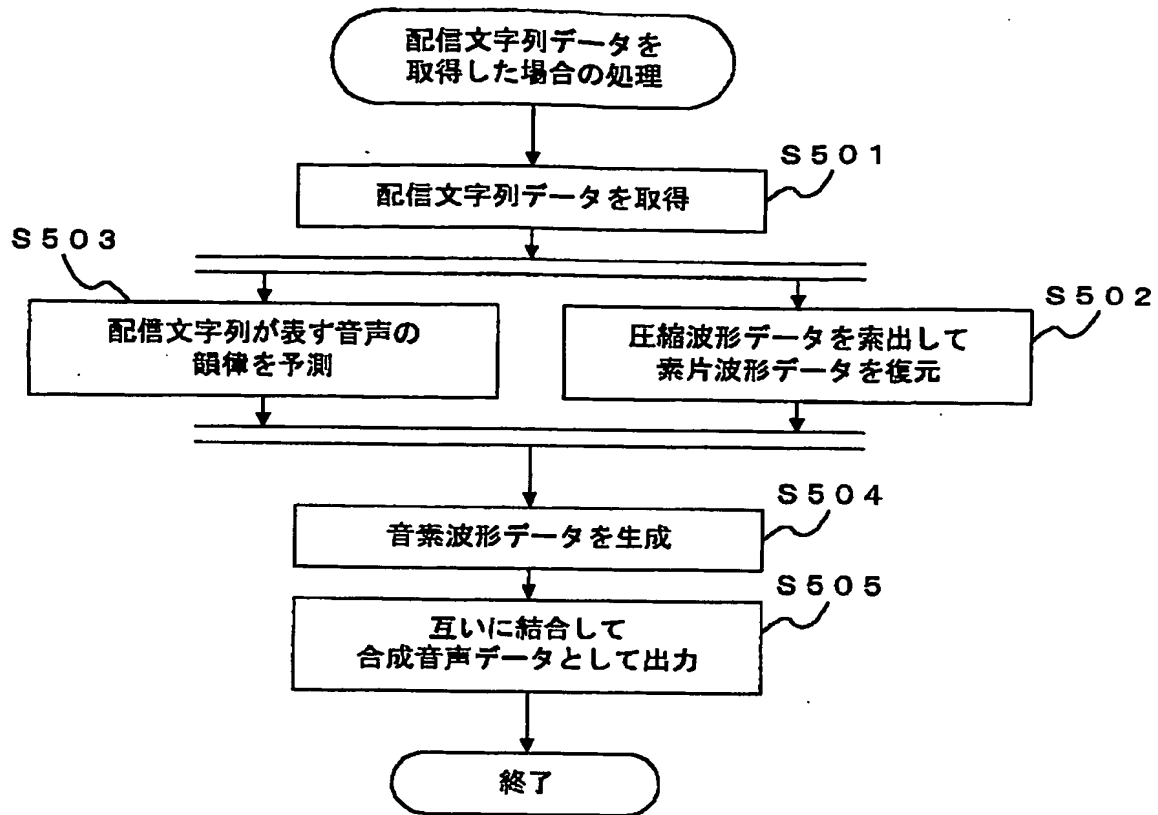
【図 6】



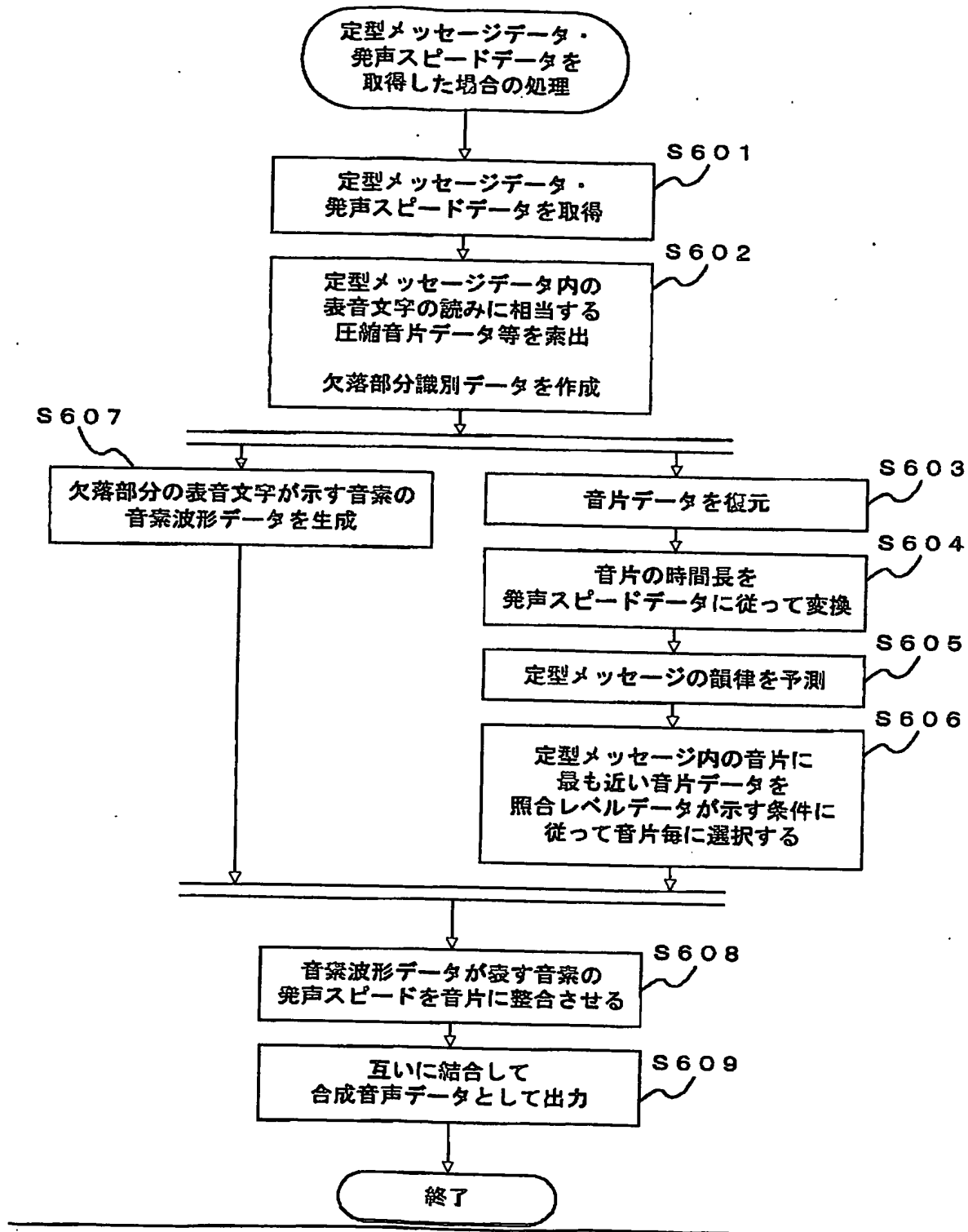
【図 7】



【図 8】



【図 9】



【書類名】 要約書

【要約】

【課題】 簡単な構成で高速に自然な合成音声を得るための音声合成装置等を提供することである。

【解決手段】 定型メッセージを表すデータが供給されると、音片編集部 5 は、定型メッセージ内の音片と読みが合致する音片の音片データを音片データベース 7 から索出させ、発声スピードデータが示すスピードに合致するよう、この音片データを変換させる。一方で音片編集部 5 は定型メッセージの韻律予測を行い、索出された音片データのうちから定型メッセージ内の各音片に最もよく合致するものを 1 個ずつ、韻律予測結果に基づいて選択する。選択できなかった音片については音響処理部 41 に単位音声毎の波形を表す波形データを供給させる。そして、選択した音片データや、音響処理部 41 に供給させた波形データを互いに結合して、合成音声を表すデータを生成する。

【選択図】 図 1

特願 2004-142907

ページ: 1/E

出 願 人 履 歴 情 報

識別番号

[000003595]

1. 変更年月日

2002年 7月26日

[変更理由]

住所変更

住 所

東京都八王子市石川町2967番地3

氏 名

株式会社ケンウッド